
The Cedars Project

Maggie Jones

The author

Maggie Jones worked at the National Library of Australia for 17 years before returning to the U.K in 1999. During her time at the NLA, Maggie became very interested in digital preservation and has worked on a number of UK based projects. She co-authored (with Neil Beagrie) *Preservation Management of Digital Materials: a Handbook* and was project manager of the Cedars project for its final year.

Abstract

Digital Preservation has become such a pressing issue for libraries world-wide that it is easy to forget what an unknown quantity it was when the Cedars project began in 1998. The Final Report from the RLG/CPA Task Force on Digital Archiving was released in 1996 and this provided a catalyst for further action within the U.K. The Cedars Project was initially funded for three years as part of the final phase of the UK eLib project. An additional year was subsequently funded to enable consolidation and wider dissemination of its findings. The Cedars project made a significant contribution to digital preservation and was able to forge good working relationships with colleagues working in related areas. It delivered both heightened awareness of digital preservation as a critical strategic issue at the broad level, as well as some specific deliverables designed to provide practical support at the operational level. Now that the Cedars project has ended, the primary responsibility within the UK for ensuring the momentum is not lost will rest with the recently formed Digital Preservation Coalition. Significant developments overseas will also feed into the global digital preservation agenda.

What was the Cedars Project?

The Cedars (CURL Exemplar in Digital Archives) project was funded as part of eLib Phase 3. It was initially a three-year project but received additional funding for a fourth and final year to enable it to consolidate its findings and disseminate them to the wider community. The project ended in April 2001, though there will be some post-Cedars activity arising from the project, as described later in this article. As the name suggests, CURL was actively involved in the project, three CURL institutions (Oxford, Cambridge and the main site, Leeds) were partners in the project with other CURL institutions serving on the Advisory Board and Management Group. UKOLN was also an active partner in the project. It also actively sought involvement and participation from a much wider range of players, both in the U.K and overseas. For example other sectors such as public libraries and archives and the British Library were involved in the project from the beginning. Publishers were also involved in the project, recognising the key role they play in the lifecycle of digital publications. At a global level, it was recognised that as related work progressed it made sense to establish close working relationships with colleagues overseas.

Why was it funded?

Before describing the work of Cedars in more detail, it is worth taking a brief historical look at the broader environment in which Cedars was funded and the rationale behind its funding. There had been growing concern (albeit by a relatively small group) that digital preservation was an important issue which needed to be addressed, particularly since 1995, when the RLG/CPA Taskforce delivered its seminal report (Watters, D. and Garrett, J. 1996). The Task Force report was used as a frame of reference for an influential workshop organised by JISC and the British Library in that same year (Long Term Preservation of Electronic Materials 1995). 1995 also saw the beginning of the The UK Electronic Libraries (eLib) programme, charged with facilitating cultural change within the UK higher education sector regarding the use of digital technology. This programme was also funded by JISC. The eLib programme, in common with many other initiatives, invested large sums into creating digital materials and encouraging the use of digital resources by institutions and their users. Unlike many other programmes, eLib recognised that this investment needed to be protected and that a natural corollary to this activity was the consequent need to ensure that these resources remained accessible into the future. Chris Rusbridge (1998) described the decision to fund a digital preservation project as part of eLib Phase 3 as follows:

"Having caused some of the increasing move of scholarly resources into the digital domain, we felt it was irresponsible to continue to ignore or to sideline preservation issues."

This quote neatly illustrates the dilemma involved in encouraging the exploitation of digital technology on the one hand, while simultaneously recognising its fragility on the other. The decision to fund Cedars as part of eLib Phase 3 was therefore a significant recognition of the high priority which JISC felt digital preservation to be as a strategic issue for UK Higher Education. Cedars accordingly focussed its efforts on the UK Higher Education Community and on scholarly resources required in support of that community.

What did Cedars achieve?

Awareness Raising. There were a number of discrete deliverables produced by Cedars and a range of general contributions. Of the latter, awareness raising was needed, not only of the importance of digital preservation per se, but crucially, of the potential role to be played by a far wider number of stakeholders than is involved in traditional preservation. While digital preservation was recognised as increasingly important, this was by a relatively small number of people. The word 'preservation' has very specific connotations for most, derived from the perception of preservation in the print environment in which preservation was the responsibility of specialists and invariably carried out, if at all, at a very much later stage from creation and acquisition. The concept of 'benign neglect' in which material may be left for decades if not hundreds of years and can still have a good chance of remaining accessible is also one which has served traditional library collections well over the years. This is a very different model to digital preservation, in which creators and owners of digital materials have a crucial role to play in their subsequent preservation. If due regard for use of standards and accompanying documentation is not given from the very beginning, it may not be possible, at least cost-effectively, to preserve the material at a later date. In the final year of the project, when the primary focus was on

consolidation and dissemination of results, the work of Cedars was categorised into five main areas and five guides were produced¹, as follows.

Cedars Guide to the Digital Archiving Prototype. An initial assumption made by the Cedars project was that digital preservation would ultimately be distributed among a number of institutions. So it was necessary to build an archive that would not only store digital materials and their associated metadata required to interpret them, but also be capable of interacting with different institutions, using different technical systems. The project adopted the Open Archival Information System (OAIS) model very early on as the framework for its archive as OAIS had already identified a standard framework for implementing a digital archive and it was pointless to duplicate effort. However the Cedars implementation also addressed external issues which were not part of OAIS, such as creators and users of digital materials. The need to negotiate with creators over issues such as copyright was felt by the Cedars team to be an essential component of a digital archive. The first phase of testing involved several institutions in testing whether digital objects could be delivered to different sites and retrieved from them. Different digital objects were used by each test site in an attempt to replicate as much as possible a real-life situation in which a digital archive must deal with many different formats. Each object was stored as a byte stream in order to liberate it from inevitable technological obsolescence but each object pointed to the technical metadata which will enable its interpretation. The decision to separate the rendering tools from the digital object itself was made in order to simplify the maintenance of the archive. The technical metadata (or representation information as it is referred to in Cedars and OAIS) can be updated relatively easily without the need to check every individual digital object which may point to it².

Cedars Guide to Digital Preservation Strategies. In the early stages of Cedars, there was much discussion in the literature of two main strategies for overcoming the inevitable technological obsolescence. These were migration and emulation and both had fierce champions and equally fierce detractors. Cedars arrived at a more subtle interpretation. The Cedars approach is to preserve both the original bytestream as well as detailed metadata which will enable it to be interpreted in the future. Preserving a bytestream is a relatively routine IT operation. Adding appropriate metadata will require new workflows at the point of acquisition. However, since classes of digital objects will be interpreted by the same technical metadata, economies of scale can quickly be achieved.

During the early stage of the Cedars project, an opportunity was taken to bid for a joint JISC/NSF funded project which could focus on researching emulation as a practical digital preservation strategy. The bid was successful and the CAMiLEON (<http://www.si.umich.edu/CAMiLEON/>) project began in 1999, a collaborative project between the universities of Leeds and Michigan. Having the UK part of the

¹ All five Cedars Guides are accessible from URL:
<http://www.leeds.ac.uk/cedars/pubconf/pubconf.html> [Accessed 12/9/02]

² Further information about the Cedars demonstrator and test sites can be found at URL:
<http://www.leeds.ac.uk/cedars/archive/testsites.html> [Accessed 12/9/02]

project based at Leeds made it particularly convenient to work collaboratively with Cedars and there has been much mutual benefit from this association. The CAMiLEON project is due to end at the end of 2002.

Cedars Guide to Intellectual Property Rights. The Cedars project recognised very early on that technical challenges are only part of the landscape and other issues, such as intellectual copyright, will need to be addressed too. Digital preservation will inevitably require copying and if permission to undertake this task for preservation purposes is not forthcoming, it may not be possible to preserve the material. This is also complicated by the fact that the copyright owner may be difficult to determine and/or there may be multiple copyright owners in a composite work.

Cedars Guide to Digital Collection Management. A major difference between digital and traditional preservation is the need to involve a much larger number of players, in particular creators and owners of the digital material. Effective lifecycle management is the best way of managing digital materials, in which the main message is to begin preservation as early in the lifecycle of a digital resource as possible, preferably at creation. Moreover it is not a one-off event, preservation management needs to accommodate inevitable technology changes over time. As universities are increasingly creating their own digital materials as well as acquiring them, they clearly have a vested interest in ensuring this significant investment is safeguarded. Regardless of which repositories assume responsibility for preserving digital materials for the long-term, it will be a whole lot easier to achieve if those materials have been responsibly managed from the earliest possible stage in their lifecycle.

Cedars Guide to Preservation Metadata. Very early on in the project, it became clear to the project team that there can be no effective digital preservation without accompanying metadata. This may seem relatively obvious now but at that time, the emphasis on metadata had been very much on its application to support resource discovery, not on what was required to preserve the resource. An outline metadata specification (Russell, K., Sergeant, D., Weinberger, E. and Day, M. 2000) was developed with the dual purpose of providing practical support for the demonstrator, but also with a wider goal of contributing to international standardisation on preservation metadata. The linkages with international developments is a recurring theme of the Cedars project and never more so than with preservation metadata. As mentioned earlier, the Cedars project made good use of the OAIS reference model, itself an internationally developed standard, and contributed to its development. The work of RLG and OCLC in attempting to arrive at a standard set of preservation metadata elements, made use of the Cedars outline specification in its deliberations. The report of the RLG/OCLC Working Group on Preservation Metadata (OCLC/RLG Working Group on Preservation Metadata 2002) brings together the experience and expertise from the international community and is an important step forward in achieving a broadly applicable standard in this area.

A Focal Point for the Wider Preservation Community. It seems appropriate to start and finish a description of Cedars' achievements with its broad role in terms of awareness raising and encouraging debate and discussion regarding further action. It was never assumed (certainly not by Cedars project staff!) that the Cedars project would provide some kind of magic solution for all the many and varied challenges of digital preservation. The importance of building and maintaining contact with

colleagues working on similar projects has been mentioned before. As well as the numerous formal and informal involvement between Cedars and the wider digital preservation community, Cedars also sought to highlight the importance of digital preservation through an international conference, Preservation 2000, held in York on 6-8 December 2000³. This event brought together more than 150 people from Australia, Europe, and North America and provided an opportunity to gain a wider perspective on digital preservation activity. As Lynne Brindley commented in her keynote address, the conference was essentially “preaching to the converted” in that those who were already very much aware of, and interested in, digital preservation attended. It still provided an excellent opportunity to bring together a range of experts and practitioners but was recognised that there would need to be additional vigorous lobbying of others, in particular funders and senior administrators. It has become axiomatic that the most effective progress is going to come through collaboration. Whether that collaboration achieves more than mere rhetoric will depend to a large extent on the active engagement of those who have sufficient grasp of the issues to be able to identify what needs to be done and with the necessary authority to organisationally and financially support it.

The final Cedars ‘event’ was a workshop held in February 2002. This was also “preaching to the converted”, though consciously so as the objective was to bring together those key players who were both the targets and potential beneficiaries of the Cedars project. The two-day workshop brought together more than 70 invited-only delegates and provided an opportunity for them to learn what Cedars has done on their behalf, to place that work in a wider context by also including sessions on related work, and last but not least, to decide what needs to happen next. It was also consciously parochial in that most of the audience were UK librarians and senior administrators. While the critical importance of seeing digital preservation as a global issue, benefiting from world-wide co-operation and collaboration, has been a cornerstone of the Cedars project, it was decided that the final workshop needed to be highly focussed on ensuring practical outcomes and identifying future priorities. Preservation 2000 had an international character, the final Cedars Workshop needed to showcase the work of the project to those it was aimed at, and who have the authority and motivation to ensure that the momentum is not lost post-Cedars. The latter was of particular concern as Cedars has provided such a readily identifiable focus of attention.

What didn’t Cedars achieve?

The final Cedars Workshop was helpful in identifying gaps needing to be filled, those for which neither Cedars nor other related projects have provided more than partial answers for. One of these was detailed information relating to the costs of digital preservation. This is both the issue most people want to know and also the one which is most difficult to provide sensible answers to as there are too many variables involved. The approach Cedars took was to identify broad cost elements associated with each aspect of digital preservation. For example costs associated with intellectual property rights could include the time taken to research IPR holders to seek necessary permission to undertake preservation actions, a cost associated with

³ Reports on Preservation 2000 and other Cedars events can be found at URL: <http://www.leeds.ac.uk/cedars/pubconf/pubconf.htm#conferencesandworkshops> [Accessed 12/9/02]

metadata is the cost of allocating detailed preservation metadata elements, and so on. What it then did was to categorise these very generally into one-off and ongoing costs and those which might be expected to increase over time (e.g., staff costs) and those which might be expected to decline over time (e.g. storage). This approach can provide a framework for calculating likely costs. However, ultimately, the message Cedars most wanted to get across was that the cost of not undertaking preservation effort will be far greater, both in terms of greatly increased costs involved in heroic rescue of material and/or in the loss of valuable scholarly resources. Related to this is the high probability that costs are more likely to reduce once factors such as economies of scale and strategic partnerships can be realised.

Further technical research was also identified, particularly into practical digital preservation strategies. The ongoing work of CAMiLEON will be particularly useful in this respect. However, despite a clear need for ongoing research, the workshop strongly reinforced the need to move beyond theory and into full-scale implementations. This will require powerful support and endorsement to ensure that appropriate funding is available.

What happens now?

In terms of specific Cedars outputs, a synthesis of the five Cedars Guides is being prepared and will be available from the Cedars website as well as in print. The Cedars website will continue to be maintained by the University of Leeds. At a broader, strategic level within the U.K, the Research Support Libraries Group will be extremely important in terms of ensuring that progress is made in scaling up the technical and organisational infrastructure capable of preserving large quantities of important digital materials. Their report, due in November 2002, is therefore awaited with keen anticipation. A range of strategic issues are being actively undertaken by the Digital Preservation Coalition (<http://www.dpconline.org>). In her keynote address to Preservation 2000, Lynne Brindley urged the conference to “raise the stakes, to make a concerted effort to get digital preservation on the agenda of key decision makers and funders in terms that they will understand and find persuasive.” This is now one of the goals and missions of the Digital Preservation Coalition. At the international level, there are some interesting indications of the importance of publisher/library collaboration for successful digital preservation. Among these, seven Mellon-funded planning projects investigated a range of issues associated with preserving scholarly e-journals⁴. Four of those projects were collaborative projects between publishers and universities. Mellon is now investigating two major approaches to preserving digital content, preserving source files (JSTOR) and presentation files (LOCKSS). Earlier this year, a joint IFLA/IPA (2002) statement in support of digital preservation was made, which recognised the importance of active collaboration between publishers and libraries, in this case, particularly national libraries. Many national libraries have already begun the challenging task of building infrastructure capable of including digital, as well as the range of materials in other formats they have traditionally assumed responsibility for. One of these, the Royal Dutch National Library has forged a close working relationship with Elsevier and they have recently announced an agreement between the two in which the Dutch National Library will essentially become a de facto national archive for Elsevier publications.

⁴ Details of the Mellon e-journal archiving projects and their reports can be found at URL: <http://www.diglib.org/preserve/ejp.htm>

In the U.K, another JISC funded initiative is conducting a risk assessment of e-journals covered by the NESLI model licences and scoping requirements for a central archive. A report covering options will be available in mid 2003. Finally, the scale of the National Digital Information Infrastructure and Preservation Program (NDIIPP) (Friedlander, A. 2002) can be expected to provide significant input to the global digital preservation community.

Conclusion

Cedars has made an important contribution to the development of an international digital preservation agenda. That agenda is moving very rapidly and has an increasing sense of urgency. No-one would deny that there is not still much to do, both in terms of research and full-scale implementations but there is no going back and there now exists a solid core of work which can be built on. A brief glimpse at a few of the developments which have occurred since the Cedars project began is enough to offer much more optimism than would have been possible four years ago.

References

Friedlander, A. (2002) The National Digital Information Infrastructure Preservation Program: Expectations, Realities, Choices and Progress to Date. *D-Lib Magazine*, April 2002. URL: <http://www.dlib.org/dlib/april02/friedlander/04friedlander.html> [Accessed 12/9/02]

IFLA/IPA (2002) *Preserving the Memory of the World in Perpetuity: a joint statement on the archiving and preserving of digital information*. URL: <http://www.ifla.org/V/press/ifla-ipa02.htm> [Accessed 12/9/02]

KB/Elsevier (2002) *National Library of the Netherlands and Elsevier Science make digital preservation history*. Press Release URL: http://www.hillandknowlton.nl/elsevierscience_kb/ [Accessed 12/9/02]

Long Term Preservation of Electronic Materials: a JISC/British Library Workshop as part of the Electronic Libraries Programme (eLib) (1995). 27-28 November at the University of Warwick.. URL: <http://www.ukoln.ac.uk/services/papers/bl/rdr6238/> [Accessed 12/9/02]

OCLC/RLG Working Group on Preservation Metadata (2002). *A Metadata Framework to Support the preservation of Digital Objects*. URL: http://www.oclc.org/research/pmwg/pm_framework.pdf [Accessed 12/9/02]

Rusbridge, C. (1998) Towards the Hybrid Library. *D-Lib Magazine*, July/August, 1998. URL: <http://www.dlib.org/dlib/july98/rusbridge/07rusbridge.html> [Accessed 12/9/02]

Russell, K., Sergeant, D., Weinberger, E. and Day, M. (2000) *Metadata for digital preservation: the Cedars project outline specification*. URL: <http://www.leeds.ac.uk/cedars/metadata.html> [Accessed 12/9/02]

Waters, D. and Garrett, J. (1996) *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access and the Research Libraries Group*, Washington, D.C, Commission on Preservation and Access. URL: <http://www.rlg.ac.uk/ArchTF/> [Accessed 12/9/02]